

## 施設トマト栽培における多面的収量予測手法の検討

上村 真史<sup>†</sup> 渡邊 真也<sup>†</sup> 堀江 達郎<sup>‡</sup> 小林 洋介<sup>†</sup> 佐藤 和彦<sup>†</sup> 岸上 順一<sup>†</sup>  
 室蘭工業大学<sup>†</sup> エア・ウォーター<sup>‡</sup>

## 1 はじめに

トマトの施設栽培において短期的な収量予測は、経営上最も重要な課題の1つである。栽培されたトマト木の画像から実の個数を検出して収量予測を行う手法 [1] も存在するが、多くの栽培施設では画像のセンシング環境が整えられておらず、外的環境とトマト木の生育状態から収量予測を行わなければならない。しかしながら、季節ごとに天候が大きく変化する日本において高精度な収量予測は容易ではない。

そこで本研究では、トマトの生育と外的環境要因との関係 [2] を考慮した多面的収量予測手法を提案する。

## 2 収集したデータ

北海道内にあるトマト施設栽培農園に設置した複合環境制御コントローラが計測した、2017年1月1日から2019年12月31日まで5分間隔で記録したデータ（以下、環境データ）を用いる。環境データは、センサ設置点の温度、湿度、CO<sub>2</sub>濃度や日射の4変数を計測地点ごとに収集した。

また、同期間に栽培されたトマトに対して人手で調査したデータ（以下、生育データ）を用いる。生育データは、果実単位での着果週や収穫週などが記録されており、収量データやトマト木の状態を表す特徴量として扱うことができる。

## 3 収量予測手法

本研究では、収量予測手法の学習アルゴリズムとして、CNN(Convolution Neural Networks)、NGBoost(Natural Gradient Boosting) の2手法を比較分析をする。比較に用いた手法を以下に簡単に説明する。

また、各手法の出力は収穫対象となり得る各週の実の収穫割合であり、予測された割合から収穫個数を算出し、評価を行う。

## 3.1 CNN

CNNは層構造に畳み込みとプーリングを追加したニューラルネットである。提案予測法では、時系列の環境データに対して、変数毎に1次元の畳み込みを行う構造を取り入れる。予測手法の学習において、環境データと生育データを直接入力として用いる。

## 3.2 NGBoost

NGBoostは勾配ブースティング決定木の一種であり、勾配の計算過程においてNatural Gradientを使用することで、確率分布の不確かさを考慮して学習を効率的に行える学習器である [3]。予測手法の学習において、環境データと生育データから生成した特徴量を入力として用いる。

## 4 実験

各手法毎に定めた入出力でトマトの収量予測モデルを作成、学習を行い評価を行う。Fig.1に実験のフロー図を示す。次に手法毎に大きく誤差が確認された週の入力データに着目して誤差の要因を考察する。

予測手法の精度指標は式(1)で示すMAPE(Mean Absoluted Percentage Error)と各週の絶対誤差率の最大値を用いて評価を行う。 $N$ はデータ数、 $y$ は実際の値、 $\hat{y}$ は予測した値である。

$$\text{MAPE} = \frac{100}{N} \sum_{i=1}^N \frac{|y_i - \hat{y}_i|}{y_i} \quad (1)$$

## 4.1 収量予測

未来のデータに対して手法の評価を行うため、トレーニングデータを2017年、2018年と2019年の25週までとし、テストデータを2019年の26週から45週までとした。なお、週番号はISO8601規格に則るものとする。学習時には、各手法毎にトレーニングデータを用いて交差検証を行い、ハイパーパラメータを決定した。

Multifaceted yield prediction approach for green house tomato

Masashi Uemura<sup>†</sup>, Shinya Watanabe<sup>†</sup>,  
 Tatsuro Horie<sup>‡</sup>, Yosuke Kobayashi<sup>†</sup>,  
 Kazuhiko Sato<sup>†</sup>, Jay Kishigami<sup>†</sup>

<sup>†</sup> Muroran Institute of Technology

<sup>‡</sup> AIR WATER INC.

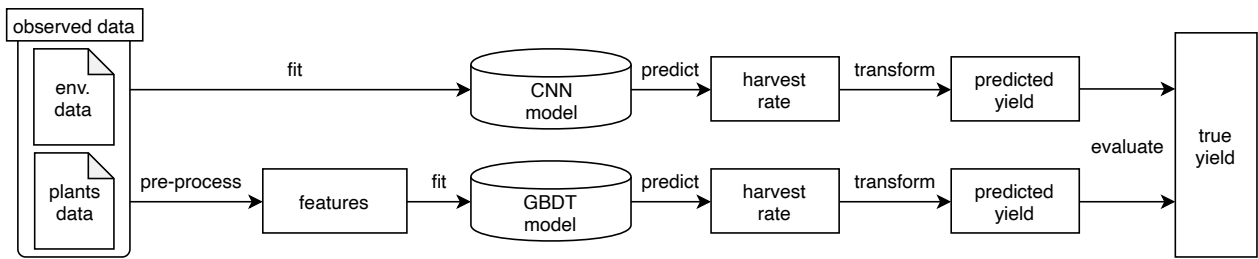


Fig.1 Flowchart of yield prediction experiment

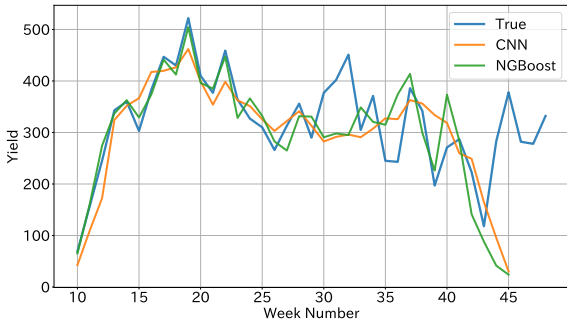


Fig.2 Results of yield prediction methods

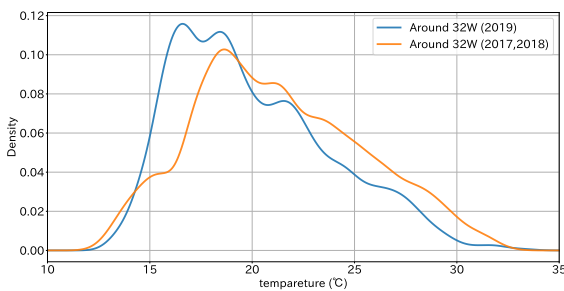


Fig.3 Kernel density estimation for each year's temperature.

Fig.2 に収量予測の実験結果を示す。横軸が週番号で、縦軸が収穫個数である。テストデータにおいてCNNはMAPEが18.36%、最大誤差が53.12%、NGBoostはMAPEが20.79%、最大誤差が47.06%となった。CNNはNGBoostに比べてMAPEが2.43%小さく精度が良いが、最大誤差ではNgboostがCNNより6.06%小さく精度が良い。これは、CNNは平均的な値を予測したことにに対し、NGBoostは収穫の時系列的ピークを捉えつつ予測を行っているためと考えられる。

#### 4.2 誤差の要因分析

本研究では、下記の2つの誤差要因に着目して要因分析を試みた。

1つ目は、年ごとの環境データの分布の違いである。Fig.3にMAPEが大きい2019年32週の入力期間と2017年、2018年同時期の気温について、カーネル密度推定を用いた確率密度関数を示す。横軸は温度、縦軸はカーネル密度である。32週はおおよそ8月上旬であり、例年では気温が高くなりや

すくトマトの成長が早まる。ピーク値を比較すると2019年は16.8度にあるのに対し、他年度は18.4度と高い値となっており、気温の分布が異なることが確認できる。このため2019年32週では、通常通り早い収穫が行われたが、学習器がその傾向を捉えることができず、誤差の発生要因となったと考えられる。したがって、年度により環境データの傾向が異なると、学習器が予測に必要な傾向を捉えることが難しくなると考えられる。

2つ目は、収量の週毎の乱高下による誤差の発生である。Fig.2において、確認できるように週毎の収穫が大きく変動している。収量予測において、この週毎の乱高下の傾向を逆に予測することが、誤差の発生要因となる。1週間後の収量を予測するような短期的な予測手法であれば、補正は前週収量と比較することで簡単に行えるが、実際の農園運用に必要な2週間後以降の収量を予測する手法の場合、乱高下の要因を究明し、その情報を有効に活用しなければならない。この乱高下の要因は、気候変化によるトマトの成長遅れに起因すると考えられるが、分析にはトマト木内部状態の生育データが必要になると考えられる。

#### 5 まとめ

本稿では、環境データと生育データからトマトの収量予測を行い、誤差の要因分析を行った。今後の課題として、誤差の要因となるデータの分析を行い、年度毎の環境データの分布の差に強いロバストな収量予測手法を検討する。

#### 参考文献

- [1] Rahnemoonfar Maryam et al., Sensors 2017, vol.17, no.905, 2017
- [2] 竹内智晴他, FIT 2018(第17回情報科学技術フォーラム), 2018
- [3] Tony Duan et al., arXiv, 2019